

Design of a Face Recognition Technique Based MTCNN and ArcFace



Dang Thai Viet, Phan Van Thien, Nguyen Huu Tu, Hoang Gia Minh,
and Ngoc-Tam Bui

Abstract The current trend of automation and data sharing in manufacturing technologies and daily living is the 4th Industrial Revolution. Computer vision technology has permeated our daily lives as a result of advancements in artificial intelligence and processing capacity. We propose using the ArcFace model, which blends deep neural networks with multi-tasking convolutional neural networks (MTCNN). The coding procedure of the agglomeration neural network facilitates the dimension-appropriate encoding of images. Techniques aimed at enhancing face recognition's most distinctive characteristics. For the face recognition model to operate at peak efficiency, the facial recognition feature must integrate with finger gestures to control smart home activities, communicate with data, and link effortlessly to smart devices via IoT technology. We construct a facial recognition model utilizing an embedded Jetson Nano computer, a fingerprint scanning module, and a Raspberry Pi camera. The IoT smart home utilizes an embedded Raspberry Pi 3B + computer. The results indicate an approximate precision of 96% and a processing speed of 16 FPS. The interface of an Internet of Things (IoT) smart house illustrates the successful execution of real-time functionalities.

Keywords Artificial intelligence (AI) · Computer vision · IoT · Smart home · Face recognition · Gesture recognition

D. T. Viet (✉) · P. Van Thien · N. H. Tu · H. G. Minh
Ha Noi University of Science and Technology, Hanoi, Vietnam
e-mail: viet.dangthai@hust.edu.vn

N.-T. Bui
Shibaura Institute of Technology, Tokyo, Japan
e-mail: tambn@shibaura-it.ac.jp

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
B. T. Long et al. (eds.), *Proceedings of the 3rd Annual International Conference on Material, Machines and Methods for Sustainable Development (MMMS2022)*, Lecture Notes in Mechanical Engineering, https://doi.org/10.1007/978-3-031-57460-3_8

1 Introduction

In the 4th Industrial Revolution, informatics and data science provided substantial support for automated production [1]. The use of this astonishing innovation resulted in a multitude of accomplishments, such as intelligent monitoring systems, sophisticated transportation infrastructure, automated financial systems and industrial assembly robot manipulators [2–5].

In keeping with this trend, we introduce the MTCNN model of face recognition for intelligent mechatronic systems. By comparing the pre-selected facial features from the image database with the person's face, the system is able to contemplate the correction of the face. Recent research, Local Binary Pattern (LBP), transformed the input image to a binary image, then partitioned the face into blocks and calculated the histogram density per block to produce the histogram feature [6]. However, the extraction of features from the histogram may be influenced by external factors such as input image quality, illumination, etc. In addition, the Dlib method correlated with HOG and SVM [7] was employed. However, there was a possibility that the accuracy would suffer if the face angle was altered. In addition, a well-known research using the FaceNet face recognition system [8] calculated the distance between face vectors using the triplet loss function. However, this technique has the problem that the quantity of math operations the computer must execute grows exponentially as the volume of input data and the overlapping features between the faces rise. To reduce this effect, the researchers employ the ArcFace model, which calculates the distance between face vectors and creates a deviation angle and an additive angular margin m to separate the characteristics of face vectors. ArcFace has developed and enhanced FaceNet, resulting in a decomposition that prevents misidentification when the original data image resembles a photo taken from a direct angle. The experimental results were reached at a frame rate of 14–16 FPS and an accuracy of approximately 96%. The findings of facial recognition for security are linked with finger gestures for home automation control. All monitoring data in real time is shown via the IoT smart home interface.

2 Methodology

The paper proposes to implement a face recognition process summarized as shown in Fig. 1.

MTCNN, an innovative algorithm for detecting faces and facial landmarks with great speed and precision, is utilized in this procedure. The MTCNN method consists of three neural networks (NN) representing three stages. In the first stage, we employ a CNN shallow to create candidate bounding boxes rapidly. The second phase refines the acquired bounding boxes using a more sophisticated CNN network. In the final step, a more advanced CNN network is used to improve the data and generate facial landmarks. Then, Arcface takes each individual's face image as input and generates

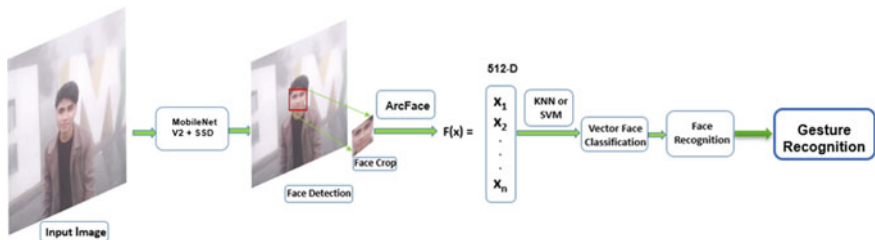


Fig. 1 Face and gesture recognition process

a vector of 512 numbers reflecting the most prominent facial traits. The term for this vector in machine learning is embedding vector. Next, a classifier is utilized to determine the distance between facial traits in order to distinguish between many identities. Due to their effectiveness in multi-class classification, Support Vector Machine (SVM) [9] and K-Nearest Neighbors (KNN) [10] are two of the most popular face recognition methods. Eventually, when the faces have been identified, users continue to run the IoT system via hand gestures.

2.1 MTCNN

The image is first rescaled to get an image pyramid that helps the model to detect faces of different sizes (Fig. 2).

2.2 ArcFace

Deep Convolutional Neural Network (DCNN) models have become prevalent for the extraction of facial features due to their exceptional benefits. There are two primary techniques to develop a classification model from vectors with facial features in order to improve the accuracy of face recognition: the triplet loss function and the softmax loss function. The softmax loss function is typically applied to situations involving face recognition [11]. The softmax loss function combines the cross entropy loss function with the softmax activation function [12]. Using the softmax function, however, causes the linear transformation matrix to grow according to the number of classes being classified. The softmax loss function L_1 is depicted here:

$$L_1 = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{W_{y_i x_i} T + b_{y_i}}}{\sum_{j=1}^n e^{W_{y_j x_j} T + b_{y_j}}} \tag{1}$$

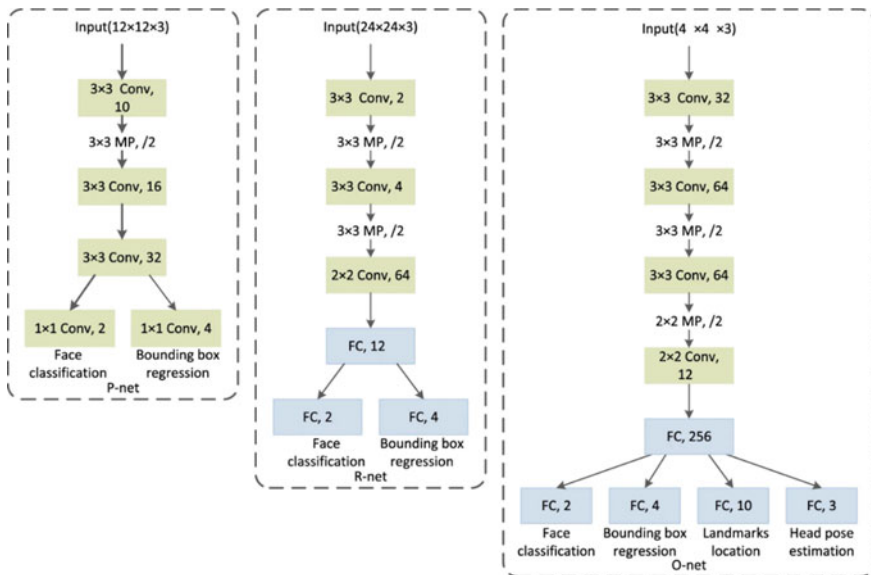


Fig. 2 MTCNN architecture

where $x_i \in R^d$ represents the depth feature of sample i , of class y_i . Embedded feature size d is set to 512. $W_j \in R^d$ represents the j th column of the weight vector $W \in R^{d \times n}$ and $b_j \in R^n$ is the bias. The batch and numeric class sizes are N and n respectively.

Since embedded features are dispersed around the center of each feature on the hypersphere, an additive angular margin penalty m is introduced between x_i and W_{y_i} , while intra-class compactness and inter-class differentiation are improved. Due to the fact that the suggested additive angular margin penalty is equivalent to the geodetic distance margin penalty in the normalized hypersphere, the method is referred to as ArcFace Lost L_2 (see Eq. 2).

$$L_2 = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s(\cos(\theta_{y_i} + m))}}{e^{s(\cos(\theta_{y_i} + m))} + \sum_{j=1, j \neq i}^n e^{s(\cos(\theta_{y_j} + m))}} \quad (2)$$

Figure 3 shows the process of training a DCNN for face recognition by the ArcFace loss function.

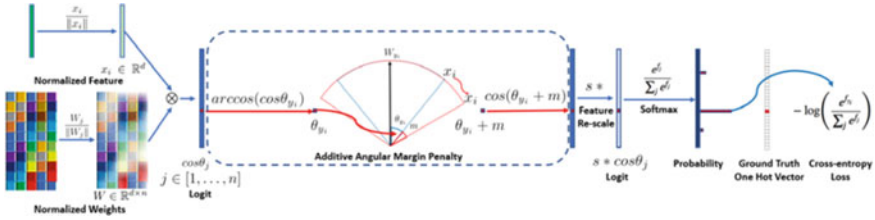


Fig. 3 Procedure for training a DCNN for recognition by ArcFace loss

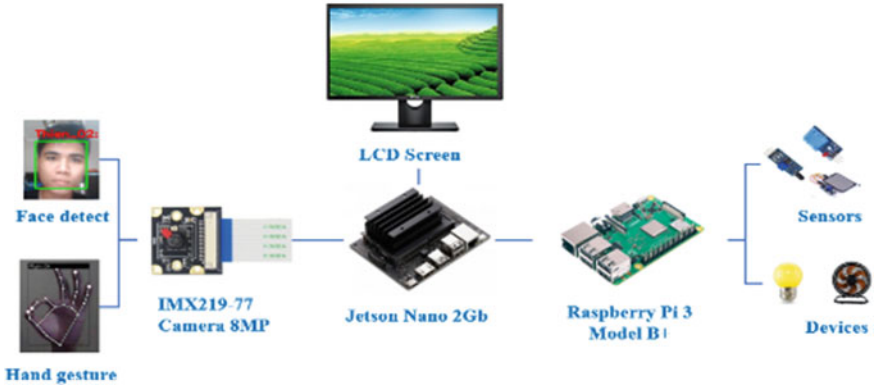


Fig. 4 Block diagram of the proposed hardware system

3 System Structure

3.1 Hardware and Software

Process and System Design

The face and hand gesture recognition system using Jetson Nano is capable of handling multiple video streams (see Fig. 4).

In this system, based on the input data set, the camera conducts face recognition in real time; if the face matches, the system will continue to allow the operator to manipulate gestures to control the IoT system (see Fig. 5).

4 Experimental Results

On training dataset with 5 frames per class. We evaluate the face recognition model using ArcFace with models such as: Dlib, LBP. The ArcFace algorithm achieves predominant performance on the Jetson Nano embedded computer with the accuracy



Fig. 5 The control interface of IoT system

of 95–97 (%) and the frame rate up to 25 (FPS). The results are depicted more clearly in Table 1.

The ArcFace model balances accuracy and speed in facial recognition. With the application of hand gesture recognition after verifying the user’s identity, the user can control the sensor devices, the light in the system is clearly displayed as being shown in Fig. 6.

Table 1 List of facial recognition test results of some models

Model	Accuracy (%)	FPS
LBP	84–88	19–21
DLIB	52–59	9–10
MTCNN + ArcFace	95–97	21–25



Fig. 6 Control features with display of biological indicators

5 Conclusions

Face recognition that improves safety and security has proven to be a formidable obstacle for researchers. We tried the application of the ArcFace model in face recognition and achieved generally favorable results. Real-time testing and evaluation with 30 distinct input faces demonstrate an inference rate of 16 FPS and an accuracy of roughly 96%. In addition, the gesture recognition function that integrates controls with a set of six operations has a 96% accuracy rate. The entire system was developed and implemented on Jetson nano, which yields the best results compared to other embedded computers (Raspberry Pi 3B+, etc.). In addition to facial recognition's security contribution in smart administration system user authentication, gesture combinations are automatically identified by hand figures, enabling automatic control and monitoring of IoT devices.

Acknowledgements This research is funded by Hanoi University of Science and Technology (HUST) under project number T2022-PC-029. The authors express grateful thankfulness to Vietnam-Japan International Institute for Science of Technology (VJIIST), School of Mechanical Engineering, HUST, Vietnam and Shibaura Institute of Technology, Japan.

References

1. R.S. Peres, X. Jia, J. Lee, K. Sun, A.W. Colombo, J. Barata, Industrial artificial intelligence in industry 4.0: systematic review, challenges and outlook. *IEEE Access* **8**, 220121–220139 (2020)
2. T.-V. Dang, N.-T. Bui, Research and design obstacle avoidance strategy for indoor autonomous mobile robot using monocular camera. *J. Adv. Transp.* (2022)
3. T.-V. Dang, N.-T. Bui, Multi-scale fully convolutional network-based semantic segmentation for mobile robot navigation. *Electronics* **12**(3), 1–18 (2023)
4. T.-V. Dang, Smart home management system with face recognition based on ArcFace model in deep convolutional neural network. *J. Robot. Control* **3**(6), 754–761 (2022)
5. T.-V. Dang, D.-S. Nguyen, Optimal navigation based on improved A* algorithm for mobile robot, in *International Conference on Intelligent Systems and Networks 2023, 18–19 March, Hanoi, Vietnam* (2023)
6. S. Pawar, V. Kithani, S. Ahuja, S. Sahu, Local binary patterns and its application to facial image analysis, in *Proceedings of the 2011 International Conference on Recent Trends in Information Technology (ICRTIT)* (2011), pp. 782–786
7. H.S. Dadi, G.K.M. Pillutla, Improved face recognition rate using HOG features and SVM classifier. *IOSR J. Electr. Commun. Eng.* **11**(4), 33–44 (2016)
8. T.-V. Dang, Smart attendance system based on improved facial recognition. *J. Robot. Control* **4**(1), 46–54 (2023)
9. M.E. Mavroforakis, S. Theodoridis, Tutorial on support vector machine (SVM) through geometry, in *Proceedings of the 2005 13th European Signal Processing Conference* (2015)
10. A. Wirdiani, P. Hridayami, A. Widiari, Face identification based on K-nearest neighbor. *Sci. J. Inform.* **6**(2), 150–159 (2019)

11. J. Wang, C. Zheng, X. Yang, L. Yang, EnhanceFace: adaptive weighted SoftMax loss for deep face recognition. *IEEE Sig. Process. Lett.* **29**, 65–69 (2022)
12. X. Li, D. Chang, T. Tian, J. Cao, Large-margin regularized Softmax cross-entropy loss. *IEEE Access* **7**, 19572–19578 (2019)